



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Identification of novel regulators of developmental hematopoiesis using Endoglin regulatory elements as molecular probes

Citation for published version:

Nasrallah, R, Fast, EM, Solaimani, P, Knezevic, K, Eliades, A, Patel, R, Thambyrajah, R, Unnikrishnan, A, Thoms, J, Beck, D, Vink, CS, Smith, A, Wong, J, Shepherd, M, Kent, D, Roychoudhuri, R, Paul, F, Klippert, J, Hammes, A, Willnow, T, Göttgens, B, Dzierzak, E, Zon, LI, Lacaud, G, Kouskoff, V & Pimanda, JE 2016, 'Identification of novel regulators of developmental hematopoiesis using Endoglin regulatory elements as molecular probes', *Blood*, vol. 128, no. 15, pp. 1928-1939. <https://doi.org/10.1182/blood-2016-02-697870>

Digital Object Identifier (DOI):

[10.1182/blood-2016-02-697870](https://doi.org/10.1182/blood-2016-02-697870)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Blood

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Identification of Novel Regulators of Developmental Hematopoiesis Using *Endoglin* Regulatory Elements as Molecular Probes

Running title: *Endoglin* GREs Target Distinct Hemogenic Precursors

Rabab Nasrallah¹, Eva Fast², Parham Solaimani^{3*}, Kathy Knezevic¹, Alexia Eliades⁴, Rahima Patel⁴, **Roshana Thambyrajah⁴**, Ashwin Unnikrishnan¹, Julie Thoms¹, Dominik Beck¹, , Chris S. Vink^{3,5}, Aileen Smith^{6**}, Jason Wong¹, **Mairi Shepherd⁶**, **David Kent⁶**, **Rahul Roychoudhuri⁷**, Fabian Paul⁸, **Julia Klippert⁸**, Annette Hammes⁸, **Thomas Willnow⁸**, Bertie Göttgens⁶, Elaine Dzierzak^{3,5}, Leonard I. Zon², George Lacaud^{4#}, Valerie Kouskoff^{4#}, John E. Pimanda^{1,9#}

¹Lowy Cancer Research Centre and the Prince of Wales Clinical School, UNSW Australia, Sydney, NSW 2052, Australia. ²Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts, USA. ³Erasmus MC Stem Cell Institute, Department of Cell Biology, Erasmus University Medical Center, 3000 CA Rotterdam, Netherlands, ⁴Cancer Research UK Manchester Institute, The University of Manchester, UK. ⁵MRC Centre for Inflammation Research, University of Edinburgh, UK. ⁶Department of Haematology, Wellcome Trust and MRC Cambridge Stem Cell Institute and Cambridge Institute for Medical Research, Cambridge University, Cambridge CB2 0XY, UK. ⁷**The Babraham Institute, Babraham Research Campus, Cambridge CB22 3AT.** ⁸Max-Delbrueck-Center for Molecular Medicine, 13125 Berlin, Germany. ⁹Department of Haematology, The Prince of Wales Hospital, Sydney, NSW 2032, Australia

*Current address- BeDataDriven B.V. Binckhorstlaan 36 M4.12, 2615 BE, The Hague, Netherlands

**Current address- GSK Pharmaceutical Company, London, UK

Correspondence:

Dr John E Pimanda: jpimanda@unsw.edu.au, +61293851003 (Tel), +61293851510 (Fax) . Dr Valerie Kouskoff: Valerie.Kouskoff@cruk.manchester.ac.uk, +44(0)1614468381 (Tel), +44(0)1614463109 (Fax). Dr Georges Lacaud: Georges.Lacaud@cruk.manchester.ac.uk, +44(0)161 446 8380 (Tel, +44(0)1614463109 (Fax)

Abstract- 200, Text- **5032**, Figures-7, References- **55**

Supplemental data: Supplemental Figures-**10**, Supplemental Tables-2, Supplemental methods

Scientific category: Hematopoiesis and Stem Cells

1 **KEY POINTS:**

- 2 • Endoglin regulatory elements target hemogenic mesoderm and hemogenic
3 endothelium
- 4 • Hemogenic progenitors can be enriched using these elements as molecular probes to
5 discover novel regulators of hematopoiesis
- 6

1 **ABSTRACT**

2 Enhancers are the primary determinants of cell identity and specific promoter/enhancer
3 combinations of Endoglin (ENG) have been shown to target blood and endothelium in the
4 embryo. Here, we generated a series of embryonic stem cell lines, each targeted with reporter
5 constructs driven by specific promoter/enhancer combinations of ENG, to evaluate their
6 discriminative potential and value as molecular probes of the corresponding transcriptome.
7 The *Eng* promoter (P) in combination with the -8/+7/+9kb enhancers, targeted cells in FLK1
8 mesoderm that were enriched for blast colony forming potential, whereas the P/-8kb enhancer
9 targeted TIE2+/c-KIT+/CD41- endothelial cells that were enriched for hematopoietic
10 potential. These fractions were isolated using reporter expression and their transcriptomes
11 profiled by RNA-seq. There was high concordance between our signatures and those from
12 embryos with defects at corresponding stages of hematopoiesis. Of the six genes that were
13 up-regulated in both hemogenic mesoderm and hemogenic endothelial fractions targeted by
14 the reporters, LRP2, a multiligand receptor, was the only gene that had not previously been
15 associated with hematopoiesis. We show that LRP2 is indeed involved in definitive
16 hematopoiesis and by doing so validate the use of reporter gene coupled enhancers as probes
17 to gain insights into transcriptional changes that facilitate cell fate transitions.

18

19

1 INTRODUCTION

2 With advances in microscopy and histology, different cell types can now readily be
3 distinguished from one another. However, the molecular characteristics that make each cell
4 type unique and help distinguish stem cells from their more differentiated progeny in a tissue,
5 are still obscure. Harvesting pure populations of stem cells is a pre-requisite to probing their
6 molecular identity. Over the years, protocols combining flow cytometry with single cell serial
7 transplantation assays have been progressively refined to purify mouse and human adult
8 hematopoietic stem cells (HSCs)^{1,2}.

9

10 One of the utilitarian benefits of determining the molecular fingerprint of a HSC is
11 that it could serve as a measurable goal when developing protocols aimed at generating HSCs
12 from differentiated cells³. The failure of current protocols to generate long-term repopulating
13 HSCs from ES/iPS cells is attributed in part to our incomplete understanding of the
14 developmental journey that mesodermal progenitors traverse in the embryo when generating
15 the complement of HSCs that are resident in the bone marrow of a newborn⁴. Determining
16 the molecular identities of embryonic HSC precursors is complicated by the lack of
17 consensus regarding the precise HSC intermediates in the embryo, functional assays that are
18 less than ideal for assessment of these intermediates and knowledge that these intermediates
19 are transitory cell populations that are present in very small numbers⁵. FLK1 expressing
20 mesodermal cells in the posterior primitive streak when isolated from the embryo and
21 cultured *in vitro* generate blast colonies that have blood, endothelial and vascular smooth
22 muscle potential⁶. Blast colony forming cell (BL-CFC) potential in FLK1+ mesoderm has
23 been estimated to be ~1:300⁷. Hemogenic potential in TIE2+c-KIT+ hemogenic endothelial
24 or VE-CAD+CD45-CD41- pre-HSC cells in the dorsal aorta that transit to hematopoietic
25 cells range from 1:100-300⁸⁻¹⁰. These functional estimates are too low to probe the molecular

1 identities of either the early hemangioblast or hemogenic endothelial cell populations in the
2 developing embryo using currently available protocols.

3
4 Cell identity is encoded within the sequences of tissue specific gene regulatory
5 elements that direct and coordinate gene expression in a cell¹¹. A number of regulatory
6 elements of hematopoietic transcription factors have previously been shown to direct reporter
7 expression to developing blood cells in the mouse embryo and include enhancers of *Scl*,
8 *Runx1*, *Gata2*, *Erg*, *Fli1*, *Lmo2* and *Lyl1*, which also form a recursive circuit in human adult
9 HSCs¹². The *Runx1*+23 enhancer marks a population of early hemogenic endothelial cells
10 that transit to HSCs and has been used to isolate cells from different embryonic stages for
11 transcriptomic analysis¹³. *Ly6a/Sca1* and Endoglin (*Eng*; CD105) serve as useful cell surface
12 markers for isolation of murine HSC fractions^{14,15}. The promoter of *Ly6a* and
13 promoter/enhancer combinations of *Eng* also target embryonic hematopoiesis and in the case
14 of the former have been used in conjunction with a reporter to isolate hemogenic endothelial
15 cells and HSCs from early embryos¹⁶⁻¹⁸.

16
17 ENG is an accessory receptor and modulator of TGF- β superfamily signaling¹⁹. ENG
18 is expressed on FLK1+ mesoderm and is required for normal BL-CFC development and its
19 expression facilitates the hematopoietic program in these cells^{10,20}. ENG null mice die at
20 E9.5 with vascular defects due to abnormal endothelial and pericyte development²¹. It is also
21 a marker of adult murine HSCs that was identified using a *Scl*+19 driven fluorescent reporter
22 coupled with transcriptomic and proteomic assessment of purified cells¹⁵. An emerging
23 concept of developmental hematopoiesis posits that HSC development from the dorsal aorta
24 at E10 reflects maturation of cells that were fated earlier during embryogenesis towards the
25 hematopoietic lineage¹³. As such we rationalised that transcriptional regulation of ENG,

1 which is functionally important for the development of hematopoietic intermediates could be
2 instructive in helping elucidate the transcriptional environment of these cells. We have
3 previously shown that sequence information within the promoter and hemato-endothelial
4 enhancers of *Eng* determine how reporter genes are targeted to either endothelial or blood and
5 endothelial tissues in the embryo^{17,22}. Given the spectrum of cell types that are involved in
6 the developmental journey of embryonic HSCs and the deterministic role that ENG plays in
7 their development, we hypothesised that distinct combinations of promoter/enhancers of this
8 gene are used by different hematopoietic intermediates to regulate *ENG* expression. We
9 rationalised that if distinct promoter/enhancer constructs indeed targeted functionally distinct
10 hematopoietic intermediates, they could be used as molecular probes to profile the
11 transcriptional environment of these cells.

12

13 Here we show using ES cells with single copy reporter coupled transgenes targeted to
14 the constitutively active HPRT locus that distinct promoter/enhancer combinations of ENG
15 are used by FLK1+ mesoderm and hemogenic endothelium that are enriched for BL-CFC and
16 hematopoietic potential respectively. Using these reporter coupled transgenes as probes to
17 harvest cell populations from ESC differentiation assays, we performed RNA-seq to identify
18 gene sets that were associated with functional enrichment of hematopoietic potential and
19 show their complementarity with primary mouse tissues at matching stages of development.
20 Of the six genes that were up-regulated in both hemogenic mesoderm and hemogenic
21 endothelial fractions targeted by the reporters, LRP2, a multiligand receptor, was the only
22 gene that had not previously been associated with hematopoiesis. Here, we show that LRP2 is
23 indeed involved in definitive hematopoiesis and by doing so validate the use of reporter gene
24 coupled enhancers as a discovery tool.

25

1 MATERIALS AND METHODS

2 Murine ES cell culture

3 The Bry/GFP⁷ and HM1 ES cells²³ were cultured as previously described. See *Supplemental*
4 *data*

5

6 Generating *Hprt* targeting constructs

7 To generate the *Hprt* targeting constructs DNA fragments corresponding to the *Eng* -8, *P*, +7,
8 +^{17,22}, *mutant* +7 (+7Δ, *mutant Gata*), and *mutant* +9 (+9Δ, *mutant Ets*) regions (Fig. S1B)
9 along with the *LacZ* reporter gene were ligated in multiple combinations into the pMP8
10 targeting vector. See *Supplemental data*

11

12 *Hprt* targeting of HM1 ES cells

13 To prepare the cells for targeting, HM1 ES cells were cultured in DMEM-ES medium
14 containing 1× 6-Thioguanine (6-TG, Sigma Aldrich) for a week before electroporation. See
15 *Supplemental data*. Genomic DNA was extracted and analysed for successful recombination
16 by PCR and southern blots.

17

18 Isolation of Genomic DNA

19 Genomic DNA isolated from recombinant ES cell clones was analysed for successful
20 integration of the DNA fragments by qPCR using primers listed in *Supplemental data*.

21

22 Southern blotting

23 10μg of genomic DNA was digested with HpaI and SacI (NEB) at 37°C O/N and blotted as
24 detailed in *Supplemental data*

25

1 **ES cell differentiation into EBs**

2 To generate EBs, ES cells were collected and cultured as detailed in *Supplemental data*. Cells
3 were then seeded on ultra-low attachment 60mm plates (Sterilin) and placed in an incubator
4 at 37°C and 5%CO₂. These culture conditions were optimal for growth of day 1-4 EBs ²⁴.

6 **LacZ staining**

7 Fluorescein di-β-D-galactopyranoside (FDG, Sigma Aldrich) was added to the cells at a 1:1
8 ratio and incubated at 37C for 1 min. The cell suspension was then added to IMDM/20%FBS
9 and placed on ice for 1 hour. Further staining with primary and secondary antibodies was
10 performed after the LacZ staining procedure. See *Supplemental data*

12 **Flow cytometry and cell sorting**

13 Cells were collected from EBs and liquid blast cultures and dissociated into a single cell
14 suspension. Details of procedure and antibodies are listed in *Supplemental data*

16 **Methylcellulose blast colony forming (BL-CFC) assay**

17 2x10⁴ FLK1+/LacZ- and FLK1+/LacZ+ cells sorted from day 3 EBs were seeded in
18 methylcellulose mix. The blast potential in Flk1+ cells was assayed in triplicate dishes of
19 1ml. The plates were then placed in a humidified incubator 37°C and 5% CO₂. Colonies were
20 scored following four days of culture²⁴. See *Supplemental data*

22 **Liquid BL-CFC assay**

23 Day 3 EBs were dissociated into single cells using trypsin (Gibco), and stained with a Flk1-
24 PE or Flk1-bio antibody (eBiosciences). Sorted FLK1+ cells were seeded on gelatin coated
25 plates at a density of 7.5-8.5x10⁴ cells/9.6cm² (in a humidified incubator 37°C, 5% CO₂). The

1 culture medium consisted of IMDM supplemented with 10% FBS pre-tested for
2 differentiation, 1% L-Glutamine, 0.6% of 30mg/ml transferrin, 0.3% of 0.15M MTG, 0.5% of
3 5mg/ml ascorbic acid, 15% D4T conditioned medium, 0.1% of 5µg/ml VEGF, and 0.1% of
4 10µg/ml IL-6 (R&D systems).

5

6 **Hematopoietic methylcellulose colony-forming assay**

7 Cells isolated from day 2 or 4 of liquid blast cultures and seeded for CFU-C assays as
8 detailed in *Supplemental data*. Primitive colonies were scored after 4-5 days of culture,
9 whereas most definitive colonies were scored after 7-10 days. All cytokines used in this assay
10 were purchased from R&D systems²⁴.

11

12 **RT-PCR**

13 List of primers and methods are in *Supplemental data*.

14

15 **Chromatin immunoprecipitation (ChIP) assay**

16 ChIP assays were performed as detailed previously²². See *Supplemental data* for a list of
17 primers and experimental details.

18

19 **Mouse embryo immunostaining and imaging**

20 10µm thick cryosections of E10.5 fixed mouse embryos embedded in a gelatin/sucrose
21 solution were thawed and rehydrated in PBS prior to immunostaining. Details of procedure
22 and antibodies are listed in *Supplemental data*.

23

24 **Generating zebrafish morpholinos and analysis**

25 Details are listed in *Supplemental data*.

1

2 **Statistical analysis**

3 RT-PCR data, BL-CFCs and hematopoietic colony counts were statistically analysed using
4 Student's T-test or Paired Student's T-test. Significant differences are marked * for $p < 0.05$,
5 ** for $p < 0.01$, *** for $p < 0.001$ and not significant differences are marked NS.

6

7 **RNA sequencing and analysis**

8 RNA was isolated from cells as per standard method (For details about samples refer to
9 supplementary table S3), and amplified using the Ovation RNA amplification system V2
10 (Nugen). Single-stranded TruSeq cDNA libraries were generated and sequenced using the
11 Illumina HiSeq2000 analyzer (BGI, Hong Kong). The data has been deposited in GEO under
12 the accession number GSE77390. See *Supplemental data* for details.

13

14 **Pathway and Gene-set enrichment analysis**

15 The Ingenuity IPA Core Analysis Tool (version 17199142) was used to establish 'Molecular
16 and Cellular Functions' that correlate with the lists of differentially expressed genes. Gene-
17 set enrichment analysis was performed using the GSEA Java Desktop tool (v 2.0.13). See
18 *Supplemental data* for details.

19

RESULTS

Mesoderm to hemangioblast transition is accompanied by increased *Eng* expression and chromatin accessibility at hemato-endothelial regulatory elements.

The promoter of *ENG* when coupled with -8kb, +7kb and +9kb enhancers have previously been shown to direct reporter expression to either endothelial or blood and endothelial tissues in the embryo (Figure 1A; ^{17,22}). The Bry-GFP ESC line has been used extensively to investigate the developmental progression of pre-mesoderm (GFP-/FLK1-) to pre-hemangioblast mesoderm (GFP+/FLK1-; G+/F-) to the hemangioblast (GFP+/FLK1+; G+F+)⁷ (Figure 1B). We used this cell line to first evaluate expression of *Eng* and chromatin accessibility at hemato-endothelial regulatory elements of *Eng*^{17,22} as cells progressed from pre-hemangioblast mesoderm to hemangioblast mesoderm. *Eng* expression increased by ~ 3-fold (Figure 1C) and enrichments of H3K9 acetylation (an active chromatin mark) increased ~ 10-20 fold at the *Eng* promoter and -8kb, +7kb and +9kb *Eng* enhancers (Figure 1D). There was no change in H3K9Ac at -4kb, a region that is highly conserved across species but shows no enhancer activity²².

The *Eng* promoter when combined with the -8, +7 and +9 hemato-endothelial enhancers, targets FLK1+ mesodermal cells enriched for BL-CFC potential.

HM1 ESCs have a disrupted *Hprt* locus that can be reconstituted by homologous recombination of a targeting vector²³. They serve as a useful tool to evaluate reporter activity of single copies of gene regulatory elements at a constitutively active locus at different stages of ESC differentiation. We took advantage of this system to introduce combinations of *Eng* regulatory elements with blood and endothelial activity in *in vivo* transgenic assays (Figure S1). Successful recombination and generation of ESC lines with -8/P/lacZ, -8/P/lacZ/+7, -8/P/lacZ/+9, -8/P/lacZ/+7/+9, -8/P/lacZ/+7Δ (GATA)/+9 and -8/P/lacZ/+7Δ (GATA)/+9 Δ

(ETS), was confirmed by RT-PCR and southern blotting (Figure S2). We used these ESC lines as a tool-kit with which to track, evaluate and compare the activity of each of these gene regulatory elements (GREs) during different stages of hematopoietic development and to fractionate cells for functional validation and transcriptomic analysis.

To identify which, if any of the *Eng* GREs, targeted FLK1+ mesoderm enriched for hemangioblast potential, we generated embryoid bodies (EBs) from each ESC line and fractionated FLK1+lacZ- (F+L-) and FLK1+lacZ+ (F+L+) cells and performed blast colony forming cell (BL-CFC) assays (Figure 2A). The *Eng* -8/P/lacZ/+7/+9 construct, which showed robust blood and endothelial staining *in vivo*¹⁷ targeted a fraction of the FLK1+ mesoderm that showed increased (~ 4 fold) BL-CFC potential (Figure 2B and Figure S3A). We have previously shown that mutating the GATA binding motifs in +7 and ETS binding motifs in +9 diminished endothelial activity and extinguished hematopoietic activity of the *Eng* -8/P/lacZ/+7/+9 construct in transgenic assays¹⁷. There was a corresponding reduction or failure of the mutant constructs to preferentially target cells with BL-CFC potential (Figure 2C (i)-(ii) and Figure S3B). The -8P/lacZ construct showed strong endothelial but no hematopoietic activity in transgenic assays¹⁷. FLK1+ cells targeted by this construct (F+L+) showed significantly lower BL-CFC potential than F+L- cells (Figure 2D (i) and Figure S3A). For the -8P/lacZ/+7 and -8P/lacZ/+9 constructs, which showed strong endothelial and low- to moderate hematopoietic activity in *in vivo* transgenic assays¹⁷, FLK1+ mesoderm (F+L+) had either lower or equivalent BL-CFC activity than FLK1+ (F+L-) cells (Figure 2D (ii)-(iii) and Figure S3A). It is important to note that the total number of BL-CFCs generated by FLK1+ mesoderm will vary from clone to clone but that comparisons are of BL-CFC potential of L+ and L- sorted FLK1+ cells from each clone.

1 Taken together, these data show that the GREs of ENG that showed increased
2 chromatin accessibility as pre-hemangioblast mesoderm progressed to hemangioblast
3 mesoderm (i.e. -8/P/+7/+9; Figure 1D) act collectively to target reporter gene expression to
4 BL-CFCs in FLK1+ mesoderm. It is noteworthy that in *in vivo* transgenic assays, it was also
5 this construct (-8/P/lacZ/+7/+9) that had the strongest and most specific activity in blood and
6 endothelium in the developing embryo¹⁷.

7
8 **Global transcriptomic analysis of FLK1 mesoderm targeted by *Eng* -8/P/LacZ/+7/+9**

9 **identifies genes associated with hemangioblast activity.** To discover genes associated with
10 the activation of these GREs and increased activity of the reporters, we performed RNA-
11 sequencing on sorted *lacZ*⁺ and *lacZ*⁻ cell fractions from three independent experiments. As
12 expected *Kdr* (*Flk1*) expression was comparable in both fractions and *Eng* transcripts were
13 increased in the *lacZ*⁺ fraction consistent with *Eng* GRE driven reporter activity (Figure 3A).
14 There was also a shared set of genes that was consistently differentially expressed between
15 the *lacZ*⁺ and *lacZ*⁻ cell fractions (Figure 3B (i)) and included 107 up-regulated and 101
16 down-regulated genes. These included cell surface receptors and transcription factors known
17 to be associated with blood and endothelium (fold change ≥ 2 and p-value < 0.5 ; Table S1 and
18 Figure S4A) and genes that have been associated with hemangioblast development in the
19 LifeMap Sciences embryonic development compendium (Figure 3B (ii)). Individually the
20 expression of many genes known to be associated with early mesoderm (e.g. *Bry/T* and
21 *Bmp4*), blood (e.g. *Gata1* and *Tal1*) and endothelial (e.g. *Foxc2* and *Etv2*) development did
22 not vary significantly between these cell fractions (Figure 3B (ii)). **Indeed, as hemangioblasts**
23 **are a sub-population of FLK1+ mesoderm with multi-lineage differentiation potential, it**
24 **would have been unusual to see significant differences in expression of individual genes that**
25 **are strongly associated with commitment to a specific lineage. However, ingenuity pathway**

analysis (IPA) revealed that when differentially expressed genes were considered as a collective, there were strong associations with blood and blood vessel development for genes in the FLK1+/lacZ+ set (Figure S4B). Consistent with these biological functions, this gene set also showed significant associations with signaling pathways that govern endothelial development and eNOS signaling (Figure S4C-F).

To investigate the *in vivo* relevance of our gene set, we used GSEA analysis to compare expression overlaps with gene expression data from FLK1+ cells in ETV2^{25,26} and LDB1²⁷ KO embryos, both of which are defective in hemangioblasts (Figure 3C). There were strong overlaps between genes expressed in FLK1+ mesoderm targeted by *ENG* - 8/P/lacZ/+7/+9 and genes expressed in hemangioblast competent WT ETV2 and LDB1 embryos compared to ETV2+/- or LDB1-/- embryos respectively. Therefore the molecular signature of BL-CFC enriched FLK1+ mesoderm that was identified using differential reporter activity of *ENG* GREs is consistent with *in vivo* functional capacity.

The *Eng* promoter in combination with the -8 endothelial enhancer targets hemogenic endothelial cells enriched for hematopoietic potential. Definitive hematopoiesis in the embryo progresses through a TIE2+/c-KIT+/CD41- hemogenic endothelial (HE) intermediate⁹. We used a cell culture system that mirrors this *in vivo* transition to investigate whether any of the reporter ESC lines preferentially targeted HE cells and whether they could be used to isolate cell fractions that were enriched for hematopoietic potential⁹. To this end, FLK1+ cells were sorted from day 3EBs and seeded into liquid blast culture media (Figure 4A). At day 2 of culture, *lacZ*+ and *lacZ*- HE cells were isolated by FACS and re-seeded into liquid blast media (LBM) for two further days followed by flowcytometry and CFU-C assays to evaluate the hematopoietic potential of each fraction. Of the reporter ESC lines, *Eng* -

8/P/*lacZ* was unique in that it was active in a fraction of HE cells that generated more TIE2-
/CD41+ and CD45+ cells after 48 hours in culture (Figure 4B (i)- (ii)) and contained almost
all CFU-C potential (Figure 4B (iii)). Whereas *Eng* -8/P/*lacZ*/+7/+9 targeted FLK1+
mesoderm with increased BL-CFC potential (Figure 2), it did not target HE cells with
increased hematopoietic potential (Figure 4C (i) - (iii)). Indeed this and each of the other
constructs, targeted HE cells that had lower hematopoietic potential (Figure 4D-E; Figure
S5). Taken together, these data showed that not only was there a specific combination of
ENG GREs that targeted HE cells but that the combination was distinct from that which
targeted BL-CFCs in FLK1+ mesoderm.

**Hematopoietic potential is highest in *Eng* -8/P/*lacZ* targeted HE cells that do not as yet
express surface ENG.** Cell fate transitions are dynamic and our purpose was to use these
reporter constructs to capture HE cells that were intrinsically fated towards the hematopoietic
lineage at the earliest possible time point in culture. Based on the assumption that there
would be a delay between transactivating the *Eng* GRE reporter and surface expression of
ENG, we repeated the experiments described in Figure 4 using the *Eng* -8/P/*lacZ* ES cell line
but here also incorporating surface ENG expression to isolate TIE2+/C-KIT+/CD41- HE
fractions that were ENG+/lacZ-, ENG+/lacZ+ or ENG-/lacZ+ (Figure 5A). Interestingly,
CFU-C potential within the *lacZ*+ fraction was highest in ENG-/lacZ+ HE cells (Figure 5B
(i) – (ii)). ENG+/lacZ-, ENG+/lacZ+ and ENG-/lacZ+ HE cells were re-seeded in LBM and
analysed by flow cytometry and CFU-C assays after two further days of culture. The
proportions of TIE2-/CD41+ (Figure 5C (i)) and CD45+ (Figure 5C (ii)) cells and CFU-C
potential (Figure 5C (iii)) were highest for cultured ENG-/lacZ+ HE cells. Taken together,
these data show that the hematopoietic potential within HE cells can be targeted by *Eng* -

8/P/*lacZ* and that these ESCs could be used to interrogate the earliest transcriptional changes associated with this cell fate decision.

Transcriptomic analysis of HE fractions identifies genes associated with hemogenic endothelial to hematopoietic transition.

To discover genes, that act on *Eng*⁻8/P/*lacZ* and drive reporter gene activity and by extension are associated with hemogenic potential in TIE2⁺/C-KIT⁺/CD41⁻ HE cells, we performed RNA-sequencing on sorted *ENG*⁺/*lacZ*⁻, *ENG*⁺/*lacZ*⁺ and *ENG*⁻/*lacZ*⁺ HE cell fractions from three independent experiments. As expected the fractions, which expressed surface *ENG* had abundant *Eng* transcripts, which were still comparatively low in *ENG*⁻/*lacZ*⁺ HE cells (Figure 6A). Consistent with its role as a major determinant of endothelial to hematopoietic transition (EHT)^{9,24,28}, *Runx1* transcripts were abundant in HE cells that were enriched with functional hemogenic cells (*ENG*⁻/*lacZ*⁺ HE) and relatively low in those (*ENG*⁺/*lacZ*⁻ and *ENG*⁺/*lacZ*⁺ HE) that were not. In total, there were 707 up-regulated and 981 down-regulated genes in *ENG*⁻/*lacZ*⁺ HE cells compared with *ENG*⁺/*lacZ*⁻ and *ENG*⁺/*lacZ*⁺ HE cells (Figure 6B (i); Table S2). It was interesting to note that only a subset of genes that have previously been attributed to mark HE cells based on cell surface protein expression were¹⁶ differentially expressed between these functionally distinct HE sub-populations (Figure 6B (ii)). This does not imply that these genes are not important but that their higher or lower expression is not associated with these early subtle transitions.

To investigate the *in vivo* relevance of our gene set, we used GSEA analysis to compare expression overlaps between *ENG*⁻/*lacZ*⁺ HE vs. *ENG*⁺/*lacZ*⁻ HE and gene sets generated from primary embryonic endothelial cell (EC), hemogenic endothelial cells (HECs) and HSCs¹⁶. Consistent with our functional data, the gene sets associated with EC to HE

1 transition (Figure 6C (i) and HE to HSC transition (Figure 6C (ii)) showed strong overlaps
2 with genes expressed in ENG-/*lacZ*⁺ HE. Gene sets associated with HIF1a and DNA
3 replication also showed strong overlaps with genes expressed in ENG-/*lacZ*⁺ HE cells
4 (Figure S6A). Genes that were UP in ENG-/*lacZ*⁺ HE compared with ENG+/*lacZ*⁻ HE cells
5 feature prominently in IPA reconstructions of gene networks governing hematopoietic
6 development (Figure S6B). Whereas genes that were UP in ENG-/*lacZ*⁺ HE cells compared
7 with either ENG+/*lacZ*⁻ HE or ENG+/*lacZ*⁺ HE cells were associated with biological
8 processes relating to blood development, genes that were DOWN in ENG-/*lacZ*⁺ HE cells
9 relative to the other two fractions were associated more with angiogenesis or vasculogenesis
10 (Figure S6C-E). Interrogation of differentially expressed transcription factors (TFs) and cell
11 surface receptors (CSRs) in the more functionally hemogenic ENG-/*lacZ*⁺ HE fraction
12 relative to the ENG+/*lacZ*⁻ HE fraction showed up-regulation of a number of TFs (e.g.
13 *Runx1*²⁹, *Myb*³⁰, *Gfi1b*³¹ etc.) and CSRs (*Lgr5*³²) that are known to play a role in HSC
14 development and down regulation of others (e.g. *Sox17*³³), which are important for HE to
15 HSC transition (Figure 6D).

16
17 **Lrp2 is required for normal blood emergence in the zebrafish aorta.** We then overlapped
18 our gene sets to visualise associations between genes that were UP or DOWN in haemogenic
19 mesoderm (HB) and/or hemogenic endothelium (HE) (Figure S7 and S8; Table S1-S2) to
20 interrogate their function. Six genes were shared between the up-regulated groups (Figure
21 7A; Figure S7C) and eight genes between the down-regulated groups (Figure S8A). Genes
22 that were DOWN in both HB and HE cells included several with no known association with
23 haematopoiesis (Figure S8B). However, we focused on genes that were UP in both HB and
24 HE cells (Figure 7A and S7C) for practical considerations given that their expression and
25 functional role would be easier to validate. This group included hematopoietic transcription

1 factors (*Gfi1*³¹ and *Lyl1*³⁴), a platelet protein kinase C substrate (*Plek*³⁵) and granulocyte
2 lysosomal and lysosomal membrane proteins (*Mpo*³⁶ and *Laptm5*³⁷), all of which have known
3 functions in the hematopoietic system. It also included a multifunctional ligand (*Lrp2*) with
4 no previously described role in blood or blood development. Lrp2/Megalin is a member of an
5 endocytic receptor complex that is involved in maternal-fetal transport of folate and other
6 nutrients, lipids and morphogens such as sonic hedgehog (Shh) and retinoids³⁸. Given these
7 associations we postulated that *Lrp2* up-regulation in blood precursors was likely to be of
8 functional significance.

9
10 The *Ly6aGFP* (Sca1) mouse model, in which all HSCs throughout development are
11 GFP+^{14,39} has facilitated the study of EHT. These mice were used to show in real-time, the
12 transition of morphologically flat endothelial GFP+ cells in the E10.5 aorta to round GFP+
13 cells that co-express other HSC markers⁴⁰. Given that LRP2 was up-regulated in HE cells,
14 we evaluated LRP2 expression in *Ly6aGFP* E10.5 AGM. LRP2 shows specific expression in
15 endothelial cells with strong expression in *Ly6aGFP*+ endothelial cells and hematopoietic
16 clusters (Figure 7B).

17
18 EHT is an evolutionarily conserved process in vertebrates and real-time imaging of
19 transgenic zebrafish embryos has also shown the transition of aortic endothelial cells to
20 hematopoietic cells^{41,42}. *Lrp2* is highly conserved across different vertebrate species (Figure
21 7C; average sequence identity across all species shown = 70 %). The zebrafish genome has
22 two closely related protein- coding genes, *lrp2a* on chr. 9 and *lrp2b* on chr. 12, both of which
23 are expressed at 24-72hpf⁴³. To validate the involvement of LRP2 in HSC generation, we
24 used a zebrafish morpholino oligo (MO) knockdown approach targeting both *lrp2a* and *2b*
25 together and each alone. At 36h post fertilization (hpf), morphants were assayed by ISH for

1 *cmyb* and *runx1*, markers for emerging blood progenitors in the aorta⁴⁴. WT embryos showed
2 robust *cmyb* and *runx1* expressing cells along the dorsal aorta in contrast to *lrp2a/b*
3 morphants that showed severe reductions (Figure 7D (i); Figure S9A). There was partial
4 rescue of AGM blood progenitors when *lrp2a/b* morphants were co-injected with *hLRP2*
5 mRNA. The partial rescue was probably due to only partial homology of protein sequences
6 between humans and fish (~ 65%) and quality of *in vitro* transcribed mRNA given the large
7 size of *LRP2* cDNA (~14kb). To exclude non-specific toxicity related loss of *cmyb* and
8 *runx1* expressing cells, we co-injected *lrp2a/b* MO with *tp53* MO and saw no restitution of
9 *cmyb* expressing cells in the morphants (Figure S9B). Injection of *lrp2b*MO but not
10 *lrp2a*MO, reduced the numbers of *cmyb* expressing AGM blood progenitors (*lrp2b*MO;
11 Figure 7D (ii); *lrp2a*; data not shown). To establish that this defect in blood cell production
12 was not a secondary to loss of vascular integrity, we injected *lrp2a/b* MO into *flk:zsgreen*
13 transgenic embryos and saw no difference between morphants and controls at 32 hpf (Figure
14 7E). In addition to vascular integrity we also assessed blood flow in morphants. Both heart
15 function as well as blood flow was indistinguishable from control embryos (data not shown).
16 Taken together these data support a role for LRP2 during AGM hematopoiesis.

17

18

19

DISCUSSION

Regulatory elements of genes that demonstrate tissue specific expression have previously been used to target and characterise various cell populations in ESC systems⁴⁵⁻⁴⁸. They have also been used in conjunction with cell surface markers to isolate cell fractions at specific stages during embryonic development for the *de novo* discovery of regulators of hematopoiesis (e.g. Runx¹³ and Ly6a (Sca1)¹⁶). These studies have in the main utilized reporter gene knock-in ESCs that disrupted a functional allele or transgenic mouse lines derived by pro-nuclear injections, which were subject to copy number and/or position effect variegation. Here we used a method to mitigate these caveats and leveraged prior knowledge that promoter/enhancer combinations of *ENG* specifically targeted endothelium or both blood and endothelium in the developing embryo^{17,22} to interrogate the transcriptomes of hemogenic progenitors.

When initiating these experiments, we did not envisage that distinct combinations of *ENG* promoter/enhancers would target haemangioblast potential in FLK1 mesoderm and hemogenic potential in TIE2+/C-KIT+/CD41- hemogenic endothelium. In retrospect, given the distinct transcriptomes and functional properties of haemangioblasts and hemogenic endothelium, this should not have come as a surprise. Nor did we predict that hemogenic potential would be enriched in *ENG*- HE1 cells targeted by the *ENG* -8*PlacZ* transgene. Given that F+L+ cells expressed higher levels of *ENG* (Figure 3A), these data raise the question whether HE1 cells emerge from F+L- *ENG* low cells which are less able than their F+L+ counterparts to generate BL-CFCs or whether F+L+ *ENG* high cells subsequently shutdown *ENG* expression to facilitate their haemogenic potential in HE1 cells. As these populations were targeted by different transgenes (F+L+; *Eng*-8/*P/lacZ*/+7/+9 and *ENG*-L+ HE; *Eng*-8/*P/lacZ*), this could not be directly tested. However, ES cells targeted with dual

1 reporters each driven by either *Eng-8/P/+7/+9* or *Eng-8/P* may assist in addressing this
2 specific question.

3
4 *Lrp2*, a gene that encodes megalin, a multiligand uptake receptor that regulates
5 circulating levels of diverse compounds⁴⁹ emerged as a novel regulator of hematopoiesis.
6 Mutations in LRP2 result in impaired neuro-epithelial development and are causative of
7 Donnai-Barrow and facio-oculo-acoustico-renal syndromes⁵⁰. It has been implicated in
8 balancing BMP4 and SHH signaling in neuro-epithelium by acting as a clearance receptor for
9 BMP4 and by concentrating or depleting SHH by ligand recycling or clearance respectively
10 in a cell type and context dependent manner⁵¹. This is of mechanistic interest as the BMP4-
11 SHH gradient between the neural tube and dorsal aorta has also been implicated in the
12 induction of the HSC developmental program in the ventral wall of the dorsal aorta⁵². On a
13 C57Bl/6N background the LRP2 mutation causes lethality in mice around the time of birth
14 and there are no mutant pups although embryo collections at all embryonic stages to E18.5
15 show expected Mendelian ratios. LRP2 mutations on 129 or CD1 backgrounds also do not
16 yield survivors (Hammes et al, unpublished data). On a FVB/N background however, LRP2
17 null mice are viable with neural tube defects and this receptor has previously been implicated
18 in folate endocytosis in the developing neural tube⁵³. However, peripheral blood and bone
19 marrow hematopoietic stem and progenitor cell numbers were comparable in FVB/N wild
20 type and mutant adult mice at 6-9 months of age (Figure S10). A more detailed analysis of
21 embryonic hematopoiesis in mutant mice on both C57Bl/6N and FVB/N backgrounds will be
22 required to establish whether the numbers of emergent HSCs differ at various time points and
23 the identity of any modifier genes in FVB/N that compensate for the loss of LRP2 and these
24 investigations are ongoing. However, taken together with the zebrafish data, which shows
25 reduction rather than loss of HSCs, LRP2 is likely to facilitate (or provide a fitness

1 advantage) rather than be absolutely required for EHT. Indeed, it is important to keep in
2 mind that *Lrp2* transcripts were higher in HB cells with greater BL-CFC potential and HE
3 cells with greater CFU-C potential but cells with lower numbers of transcripts were also able
4 to generate BL-CFCs and CFU-Cs.

5
6 Deficiency of dietary folate also results in impaired neural tube development and
7 megaloblastic anemia⁵⁴. Targeted inactivation of the reduced folate carrier (RFC1), which
8 facilitates folate delivery into cells results in embryonic lethality at E10.5 due to neural and
9 hematopoietic defects⁵⁵ and components of the Megalin complex are amongst the most
10 significantly disrupted genes in null embryos³⁸. Coordinated up-regulation of a receptor that
11 facilitates folate uptake in hemogenic endothelial cells would be consistent with demand for
12 an essential hematinic in cells that are on the threshold of a replicative phase.

13
14 Although we focused our attention on *Lrp2*, as a gene without a described role in
15 hematopoiesis, from a list of six that were up-regulated in both hemangioblasts and HE cells,
16 there were other genes that were UP in only one or the other cell fraction. Given the overlap
17 of these gene sets with those generated from gene knockout embryos that showed stage
18 specific developmental defects or from sorted primary cells harvested from embryos, they
19 will serve as a rich resource to explore and manipulate the emergence of hemangioblasts from
20 FLK1+ mesoderm or definitive hematopoiesis in HE cells. Insights gained from these
21 manipulations will in turn inform tissue regeneration protocols that aim to generate functional
22 HSCs.

1 **ACKNOWLEDGEMENTS**

2 This work was supported by grants from the National Health and Medical Research Council
3 of Australia, Australian Research Council and the Dr Tom Bee Stem Cell Research Fund to
4 JEP, Cancer Research UK to VK and GL, the BBSRC, Leukaemia and Lymphoma
5 Research, The Leukaemia and Lymphoma Society, Cancer Research UK, and core support
6 grants by the Wellcome Trust to the Cambridge Institute for Medical Research and Wellcome
7 Trust - MRC Cambridge Stem Cell Institute to BG.

8

9 **AUTHOR CONTRIBUTIONS**

10 R.N, E.F, P.S, K.K, **R.T., M.S., J.K.**, A.U, J.T, D.B, A.E and R.P performed research and
11 analysed data. A.S, J.W, A.H, **D.K., R.R., T.W.**, B.G, E.D, and L.Z contributed essential
12 reagents and advice with data analysis and interpretation. R.N, G.L, V.K and J.E.P
13 contributed to study design, data interpretation and manuscript preparation. All authors have
14 read and approve the manuscript.

15

16 **DISCLOSURES**

17 No conflicts of interest to declare.

18

REFERENCES

1. Kiel MJ, Yilmaz OH, Iwashita T, Terhorst C, Morrison SJ. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell*. 2005;121(7):1109-1121.
2. Notta F, Doulatov S, Laurenti E, Poepl A, Jurisica I, Dick JE. Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science*. 2011;333(6039):218-221.
3. Riddell J, Gazit R, Garrison BS, et al. Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors. *Cell*. 2014;157(3):549-564.
4. Pereira CF, Lemischka IR, Moore K. 'From blood to blood': de-differentiation of hematopoietic progenitors to stem cells. *EMBO J*. 2014;33(14):1511-1513.
5. Medvinsky A, Rybtsov S, Taoudi S. Embryonic origin of the adult hematopoietic system: advances and questions. *Development*. 2011;138(6):1017-1031.
6. Huber TL, Kouskoff V, Fehling HJ, Palis J, Keller G. Haemangioblast commitment is initiated in the primitive streak of the mouse embryo. *Nature*. 2004;432(7017):625-630.
7. Fehling HJ, Lacaud G, Kubo A, et al. Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. *Development*. 2003;130(17):4217-4227.
8. Rybtsov S, Sobiesiak M, Taoudi S, et al. Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region. *J Exp Med*. 2011;208(6):1305-1315.
9. Lancrin C, Sroczynska P, Stephenson C, Allen T, Kouskoff V, Lacaud G. The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. *Nature*. 2009;457(7231):892-895.
10. Nasrallah R, Knezevic K, Thai T, et al. Endoglin potentiates nitric oxide synthesis to enhance definitive hematopoiesis. *Biol Open*. 2015;4(7):819-829.
11. Davidson EH. The regulatory genome : gene regulatory networks in development and evolution. Amsterdam ; London: Elsevier/Academic Press; 2006.
12. Beck D, Thoms JA, Perera D, et al. Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. *Blood*. 2013;122(14):e12-22.
13. Swiers G, Baumann C, O'Rourke J, et al. Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. *Nat Commun*. 2013;4:2924.
14. Ma X, Robin C, Ottersbach K, Dzierzak E. The Ly-6A (Sca-1) GFP transgene is expressed in all adult mouse hematopoietic stem cells. *Stem Cells*. 2002;20(6):514-521.
15. Chen CZ, Li M, de Graaf D, et al. Identification of endoglin as a functional marker that defines long-term repopulating hematopoietic stem cells. *Proc Natl Acad Sci U S A*. 2002;99(24):15468-15473.
16. Solaimani Kartalaei P, Yamada-Inagawa T, Vink CS, et al. Whole-transcriptome analysis of endothelial to hematopoietic stem cell transition reveals a requirement for Gpr56 in HSC generation. *J Exp Med*. 2015;212(1):93-106.
17. Pimanda JE, Chan WY, Wilson NK, et al. Endoglin expression in blood and endothelium is differentially regulated by modular assembly of the Ets/Gata hemangioblast code. *Blood*. 2008;112(12):4512-4522.
18. Pimanda JE, Gottgens B. Gene regulatory networks governing haematopoietic stem cell development and identity. *Int J Dev Biol*. 2010;54(6-7):1201-1211.
19. Cheifetz S, Bellon T, Cales C, et al. Endoglin is a component of the transforming growth factor-beta receptor system in human endothelial cells. *J Biol Chem*. 1992;267(27):19027-19030.
20. Perlingeiro RC. Endoglin is required for hemangioblast and early hematopoietic development. *Development*. 2007;134(16):3041-3048.
21. Li DY, Sorensen LK, Brooke BS, et al. Defective angiogenesis in mice lacking endoglin. *Science*. 1999;284(5419):1534-1537.
22. Pimanda JE, Chan WY, Donaldson IJ, Bowen M, Green AR, Gottgens B. Endoglin expression in the endothelium is regulated by Fli-1, Erg, and Elf-1 acting on the promoter and a -8-kb enhancer. *Blood*. 2006;107(12):4737-4745.
23. Magin TM, McWhir J, Melton DW. A new mouse embryonic stem cell line with good germ line contribution and gene targeting frequency. *Nucleic Acids Res*. 1992;20(14):3795-3796.
24. Sroczynska P, Lancrin C, Kouskoff V, Lacaud G. The differential activities of Runx1 promoters define milestones during embryonic hematopoiesis. *Blood*. 2009;114(26):5279-5289.
25. Wareing S, Mazan A, Pearson S, Gottgens B, Lacaud G, Kouskoff V. The Flk1-Cre-mediated deletion of ETV2 defines its narrow temporal requirement during embryonic hematopoietic development. *Stem Cells*. 2012;30(7):1521-1531.
26. Kataoka H, Hayashi M, Nakagawa R, et al. Etv2/ER71 induces vascular mesoderm from Flk1+PDGFRalpha+ primitive mesoderm. *Blood*. 2011;118(26):6975-6986.

- 1 27. Mylona A, Andrieu-Soler C, Thongjuea S, et al. Genome-wide analysis shows that Ldb1 controls
2 essential hematopoietic genes/pathways in mouse early development and reveals novel players in hematopoiesis.
3 *Blood*. 2013;121(15):2902-2913.
- 4 28. Chen MJ, Yokomizo T, Zeigler BM, Dzierzak E, Speck NA. Runx1 is required for the endothelial to
5 haematopoietic cell transition but not thereafter. *Nature*. 2009;457(7231):887-891.
- 6 29. Okuda T, van Deursen J, Hiebert SW, Grosveld G, Downing JR. AML1, the target of multiple
7 chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell*.
8 1996;84(2):321-330.
- 9 30. Mucenski ML, McLain K, Kier AB, et al. A functional c-myb gene is required for normal murine fetal
10 hepatic hematopoiesis. *Cell*. 1991;65(4):677-689.
- 11 31. Lancrin C, Mazan M, Stefanska M, et al. GFI1 and GFI1B control the loss of endothelial identity of
12 hemogenic endothelium during hematopoietic commitment. *Blood*. 2012;120(2):314-322.
- 13 32. Liu D, He XC, Qian P, et al. Leucine-rich repeat-containing G-protein-coupled Receptor 5 marks short-
14 term hematopoietic stem and progenitor cells during mouse embryonic development. *J Biol Chem*.
15 2014;289(34):23809-23816.
- 16 33. Lizama CO, Hawkins JS, Schmitt CE, et al. Repression of arterial genes in hemogenic endothelium is
17 sufficient for haematopoietic fate acquisition. *Nat Commun*. 2015;6:7739.
- 18 34. Capron C, Lecluse Y, Kaushik AL, et al. The SCL relative LYL-1 is required for fetal and adult
19 hematopoietic stem cell function and B-cell differentiation. *Blood*. 2006;107(12):4678-4686.
- 20 35. Wehrspaun CC, Haerty W, Ponting CP. Microglia recapitulate a hematopoietic master regulator
21 network in the aging human frontal cortex. *Neurobiol Aging*. 2015;36(8):2443 e2449-2443 e2420.
- 22 36. Le Guyader D, Redd MJ, Colucci-Guyon E, et al. Origins and unconventional behavior of neutrophils
23 in developing zebrafish. *Blood*. 2008;111(1):132-141.
- 24 37. Adra CN, Zhu S, Ko JL, et al. LAPTM5: a novel lysosomal-associated multispinning membrane
25 protein preferentially expressed in hematopoietic cells. *Genomics*. 1996;35(2):328-337.
- 26 38. Gelineau-van Waes J, Maddox JR, Smith LM, et al. Microarray analysis of E9.5 reduced folate carrier
27 (RFC1; Slc19a1) knockout embryos reveals altered expression of genes in the cubilin-megalin multiligand
28 endocytic receptor complex. *BMC Genomics*. 2008;9:156.
- 29 39. de Bruijn MF, Ma X, Robin C, Ottersbach K, Sanchez MJ, Dzierzak E. Hematopoietic stem cells
30 localize to the endothelial cell layer in the midgestation mouse aorta. *Immunity*. 2002;16(5):673-683.
- 31 40. Boisset JC, van Cappellen W, Andrieu-Soler C, Galjart N, Dzierzak E, Robin C. In vivo imaging of
32 haematopoietic cells emerging from the mouse aortic endothelium. *Nature*. 2010;464(7285):116-120.
- 33 41. Bertrand JY, Chi NC, Santoso B, Teng S, Stainier DY, Traver D. Haematopoietic stem cells derive
34 directly from aortic endothelium during development. *Nature*. 2010;464(7285):108-111.
- 35 42. Kissa K, Herbomel P. Blood stem cells emerge from aortic endothelium by a novel type of cell
36 transition. *Nature*. 2010;464(7285):112-115.
- 37 43. Kur E, Christa A, Veth KN, et al. Loss of Lrp2 in zebrafish disrupts pronephric tubular clearance but
38 not forebrain development. *Dev Dyn*. 2011;240(6):1567-1577.
- 39 44. Jing L, Zon LI. Zebrafish as a model for normal and malignant hematopoiesis. *Dis Model Mech*.
40 2011;4(4):433-438.
- 41 45. Ferreira R, Spensberger D, Silber Y, et al. Impaired in vitro erythropoiesis following deletion of the Scl
42 (Tal1) +40 enhancer is largely compensated for in vivo despite a significant reduction in expression. *Mol Cell*
43 *Biol*. 2013;33(6):1254-1266.
- 44 46. Wu SM, Fujiwara Y, Cibulsky SM, et al. Developmental origin of a bipotential myocardial and smooth
45 muscle cell precursor in the mammalian heart. *Cell*. 2006;127(6):1137-1150.
- 46 47. Holtzinger A, Rosenfeld GE, Evans T. Gata4 directs development of cardiac-inducing endoderm from
47 ES cells. *Dev Biol*. 2010;337(1):63-73.
- 48 48. Feng G, Mellor RH, Bernstein M, et al. Imaging neuronal subsets in transgenic mice expressing
49 multiple spectral variants of GFP. *Neuron*. 2000;28(1):41-51.
- 50 49. Saito A, Pietromonaco S, Loo AK, Farquhar MG. Complete cloning and sequencing of rat
51 gp330/"megalin," a distinctive member of the low density lipoprotein receptor gene family. *Proc Natl Acad Sci*
52 *U S A*. 1994;91(21):9725-9729.
- 53 50. Kantarci S, Al-Gazali L, Hill RS, et al. Mutations in LRP2, which encodes the multiligand receptor
54 megalin, cause Donnai-Barrow and facio-oculo-acoustico-renal syndromes. *Nat Genet*. 2007;39(8):957-959.
- 55 51. Christ A, Christa A, Klippert J, et al. LRP2 Acts as SHH Clearance Receptor to Protect the Retinal
56 Margin from Mitogenic Stimuli. *Dev Cell*. 2015;35(1):36-48.
- 57 52. Wilkinson RN, Pouget C, Gering M, et al. Hedgehog and Bmp polarize hematopoietic stem cell
58 emergence in the zebrafish dorsal aorta. *Dev Cell*. 2009;16(6):909-916.
- 59 53. Kur E, Mecklenburg N, Cabrera RM, Willnow TE, Hammes A. LRP2 mediates folate uptake in the
60 developing neural tube. *J Cell Sci*. 2014;127(Pt 10):2261-2268.

- 1 54. Bills ND, Koury MJ, Clifford AJ, Dessypris EN. Ineffective hematopoiesis in folate-deficient mice.
2 *Blood*. 1992;79(9):2273-2280.
- 3 55. Gelineau-van Waes J, Heller S, Bauer LK, et al. Embryonic development in the reduced folate carrier
4 knockout mouse is modulated by maternal folate supplementation. *Birth Defects Res A Clin Mol Teratol*.
5 2008;82(7):494-507.
- 6
- 7

FIGURE LEGENDS

FIGURE 1. Mesoderm to hemangioblast transition is accompanied by increased *Eng* expression and chromatin accessibility at hemato-endothelial regulatory elements. (A)

Schematic representation of the *ENG* locus. The transcription start site is marked with an arrow. The -8kb, +7kb and +9kb enhancers and the promoter (P) are marked in orange, exons are marked in brown and the 5'UTR in cyan. (B) Schematic representation of *Bry*-GFP ES cell differentiation. At day 3 of EB differentiation, *Bry*-GFP+/FLK1- (G+F-) and *Bry*-GFP+/FLK1+ (G+F+) cells were sorted and analysed by RT-PCR and ChIP. (C) Bar graph showing *Eng* mRNA expression levels in sorted FLK1+ve and -ve mesodermal cell populations in day 3 EBs generated from *Bry*-GFP ES cells. (D) A bar graph showing levels of enrichment of the active chromatin mark, H3K9Ac at *Eng* -8, P, +7 and +9 hemato-endothelial enhancers relative to IgG in pre-hemangioblast mesoderm (G+F-; black) and in hemangioblast mesoderm (G+F+; gray). *Eng* -4 was included as a negative control region. **, P < 0.01, ***, p < 0.001

FIGURE 2. The *Eng* promoter when combined with the -8, +7 and +9 haemato-endothelial enhancers targets FLK1+ mesodermal cells enriched for BL-CFC potential.

(A) Schematic representation of the experimental procedure. The *Eng* -8/P/LacZ, *Eng* -8/P/LacZ/+7, *Eng* -8/P/LacZ/+9, *Eng* -8/P/LacZ/+7/+9, *Eng* -8/P/LacZ/+7Δ/+9, and *Eng* -8/P/LacZ/+7Δ/+9Δ reporter constructs were introduced by homologous recombination into the HPRT locus of HM1 ES cells. Recombinant clones were differentiated into day 3 EBs and stained for FLK1 expression and β-galactosidase activity. FLK1+/LacZ- (F+/L-; gray) and FLK1+/LacZ+ (F+/L+; blue) cells were sorted and seeded into BL-CFC assays. Fractions sorted from the *Eng* -8/P/LacZ/+7/+9 were further analysed by RNA-sequencing. (B) Flow cytometry profiles of *Eng* -8/P/LacZ/+7/+9 day 3 EBs (left). BL-CFCs from sorted F+/L-

1 (gray) and F+/L+ (blue) fractions. **(C)** (i) Flow cytometry profile of day 3 EBs derived from
2 ES cells targeted with *Eng* -8/P/LacZ/+7Δ /+9 (mutated GATA motifs in the +7 enhancer) is
3 shown to the left with corresponding BL-CFCs from sorted F+/L- (gray) and F+/L+ (blue)
4 fractions shown to the right. (ii) Flow cytometry profile of day 3 EBs derived from ES cells
5 targeted with *Eng* -8/P/LacZ/+7Δ /+9Δ (mutated GATA motifs in the +7 enhancer and
6 mutated ETS motifs in the +9 enhancer) and corresponding BL-CFCs from sorted F+/L-
7 (gray) and F+/L+ (blue) fractions. **(D)** Flow cytometry profiles of day 3 EBs and BL-CFCs
8 from sorted F+/L- (gray) and F+/L+ (blue) fractions are shown for ES cells targeted with (i)
9 *Eng* -8/P/LacZ (ii) *Eng* -8/P/LacZ/+9 and (iii) *Eng* -8/P/LacZ/+7. BL-CFC counts are the
10 total number of blast colonies generated from 2x10⁴ seeded cells. Statistical analysis was
11 using students T-test, *p<0.05, **p< 0.01.

12

13 **FIGURE 3. RNA-sequencing of FLK1 mesoderm targeted by *Eng* -8/P/LacZ/+7/+9**
14 **identifies genes associated with hemangioblast activity.** **(A)** RNA-sequencing profiles
15 showing *Kdr* (*Flk1*) transcripts (top panel) and *Eng* transcripts (bottom panel) in the F+/L-
16 and F+/L+ fractions. FPKM expression values are shown to the right. **(B)** (i) Heat map
17 representation of up- and down-regulated genes in FLK1+/LacZ- (F+/L-) and FLK1+/LacZ+
18 (F+/L+) fractions in three independent experiments. (ii) Expression (FPKM values) levels of
19 genes that have previously been associated with hemangioblast function. The left panel
20 shows a subset of genes that are differentially expressed between F+/L- and F+/L+ fractions
21 and the right panel shows a subset of genes that are not. **(C)** GSEA profiles showing the
22 correspondence of genes that are differentially expressed between F+/L- and F+/L+ fractions
23 and those that are differentially expressed in ETV2+/- vs. ETV2-/- (top panel) and LDB wt
24 vs. LDB -/- gene sets. DEG; differentially expressed genes.

25

1 **FIGURE 4. The *Eng* promoter in combination with the -8 endothelial enhancer targets**
2 **hemogenic endothelial cells enriched for hematopoietic potential. (A)** Schematic diagram
3 outlining the experimental procedure. Recombinant ES cells generated using the *Eng*
4 reporter constructs were differentiated into day 3 EBs. FLK1+ mesodermal cells were sorted
5 from representative clones for each recombinant ES cell line and cultured in liquid blast
6 media. At 48 hours, CD41-/TIE2+/c-KIT+ (HE) cells were sorted into *lacZ*+ and *lacZ*-
7 fractions. The sorted fractions were re-cultured in liquid blast media for a further 48 hours
8 followed by flow cytometry and CFU-C assays. **(B)** (i) CD41 and TIE2 expression in sorted
9 **c-KIT+** HE *LacZ*- (white) and *LacZ*+ (blue) fractions (after 2 days of re-culture) derived
10 from *Eng* -8/*P/LacZ* ES cells. (ii) Percentage of CD45+ cells generated from *LacZ*- and
11 *LacZ*+ HE fractions. (iii) Bar chart showing the number and type of hematopoietic colonies
12 generated by each fraction. **(C)** (i)-(iii) Corresponding data to (B) generated from *Eng* -
13 8/*P/LacZ*+7/+9 ES cells. **(D)** (i)-(iii) Corresponding data to (B) generated from *Eng* -
14 8/*P/LacZ*+7 ES cells. **(E)** (i)-(iii) Corresponding data to (B) generated from *Eng* -
15 8/*P/LacZ*+9 ES cells. Primitive and definitive colonies were scored after four and nine days
16 respectively. Statistical analysis was using students T-test, *, $p < 0.05$, **, $p < 0.01$.

17
18 **FIGURE 5. Hematopoietic potential is highest in *Eng* -8/*P/lacZ* targeted HE cells that**
19 **do not express surface ENG. (A)** Schematic diagram outlining the experimental procedure.
20 FLK1+ mesodermal cells were sorted from day 3 EBs generated from the *Eng* -8/*P/LacZ*
21 recombinant ES cell line and cultured in liquid blast culture media. At 48 hours, CD41-
22 /TIE2+/c-KIT+ (HE) cells were sorted into ENG+/LacZ-, ENG+/LacZ+, and ENG-/LacZ+
23 fractions. These fractions were either directly seeded into CFU-C assays (B) or re-cultured in
24 liquid blast media for a further 48 hours and analysed by flow cytometry and CFU-C assays
25 (C). **(B)** (i)- (ii) Flow cytometry to show the frequencies of CD41-/TIE2+/c-KIT+ (HE) cells

1 in *ENG*^{+/lacZ}+, *ENG*^{+/lacZ}- and *ENG*-*lacZ*⁺ fractions. (iii) CFU-C potential of each sorted
2 fractions in (i). **(C)** (i) Flow cytometry analysis of CD41 and TIE2 expression in sorted HE
3 cell fractions after two days of re-culture in liquid blast media. (ii) Bar chart showing the
4 percentage of CD45 positive cells in sorted fraction. (iii) Bar chart showing hematopoietic
5 colony numbers from each fraction. Primitive and definitive colonies were scored after four
6 and nine days respectively. Statistical analysis was using student T-test, *; $p < 0.05$, **; p
7 < 0.01 and ***; $p < 0.001$.

8

9 **FIGURE 6. Transcriptomic analysis of HE fractions identifies genes associated with**
10 **hemogenic endothelial to hematopoietic transition.** **(A)** RNA-sequencing profiles showing
11 *Eng* transcripts (top panel) and *Runx1* transcripts (bottom panel) in the E+/L-, E+/L+ and E-
12 /L+ fractions. FPKM expression values are shown to the right. **(B)** (i) Heat map
13 representation of up- and down-regulated genes in *ENG*^{+/LacZ}- (E+/L-) HE, *ENG*^{+/LacZ}
14 (E+/L+) HE and *ENG*^{-/LacZ} (E-/L+) HE fractions in three independent experiments. (ii)
15 Expression (FPKM values) levels of genes that have previously been associated with
16 hemogenic endothelium. The top panel shows a subset of genes that are differentially
17 expressed between E+/L-, E+/L+ and E-/L+ fractions and the bottom panel shows a subset of
18 genes that are not. **(C)** GSEA profiles showing the correspondence of genes that are
19 differentially expressed between the E+/L-, E+/L+ and E-/L+ fractions and those that are
20 differentially expressed in endothelial cells (EC) vs. hemogenic endothelial cells (HEC) (top
21 panel) and HECs vs. hematopoietic stem cells (HSC) gene sets. **(D)** Transcription factors and
22 cell surface receptors that are up- and down regulated in the *Eng*^{-8/P} E-/L+ HE fraction. The
23 log fold changes (logFC) and log false discovery rates (logFDR) are listed for each gene.
24 DEG; differentially expressed genes.

25

1 **FIGURE 7. Lrp2 is required for normal definitive hematopoiesis.** (A) Venn diagram
2 showing the overlap of genes that are UP in FLK1 mesoderm enriched for BL-CFCs and/or
3 HE cells enriched for hemogenic potential. (B) Immunohistochemistry of
4 E10.5 *Ly6aGFP* AGM shows co-expression of GFP and LRP2 in endothelial cells and
5 hematopoietic clusters. The insets show the same sections at low magnification. (C)
6 Homology relationships of zebrafish *lrp2a* and *lrp2b* coding sequences with that of *Lrp2* in
7 different vertebrate species. (D) ISH for the HSC marker *cmyb* in zebrafish at 36 hpf. (i)
8 Low (left-side panels) and high (right-side panels) magnification images of control zebrafish
9 (top row), *lrp2* a/b morpholinos (middle row) and *lrp2a/b* morpholinos co-injected with
10 *hLRP2* mRNA (bottom row) zebrafish. (ii) Low (left-side panels) and high (right-side panels)
11 magnification images of control zebrafish (top row) and *lrp2b* morpholinos. (E) Confocal
12 images of *flk:zsgreen* reporter embryos show an intact vasculature in both control (upper
13 panel) and *lrp2a/b* morphant (lower panel) embryos. DA; dorsal aorta, NC; notochord, NT;
14 neural tube.

15

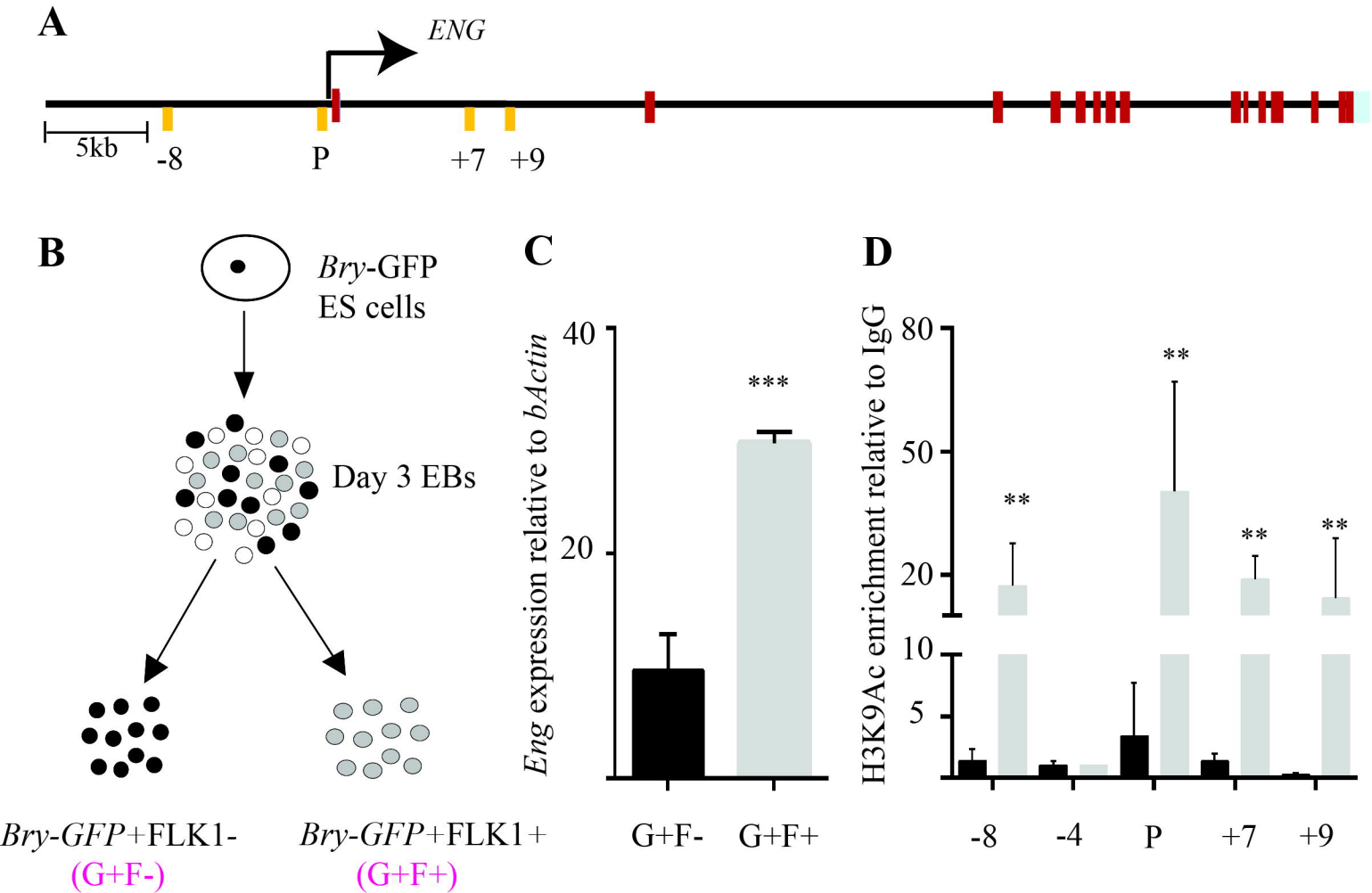
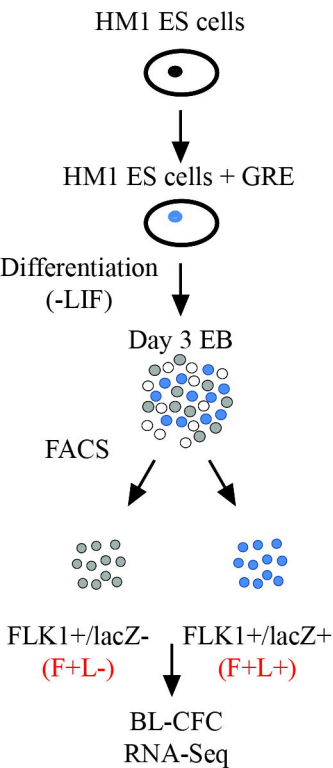
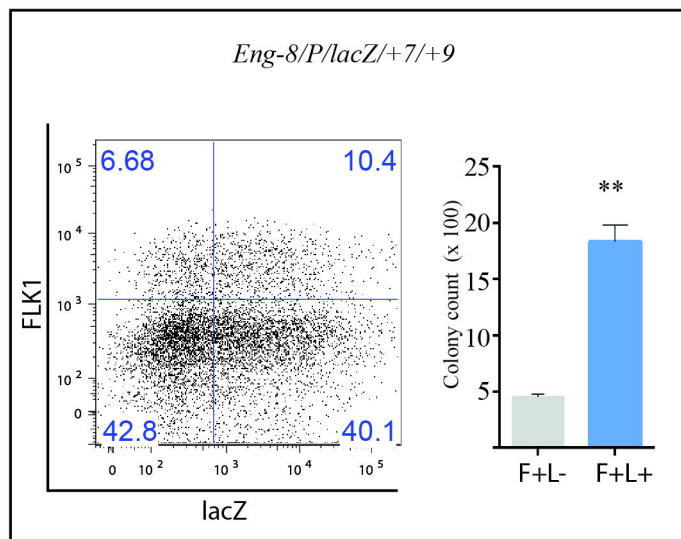


Figure 1

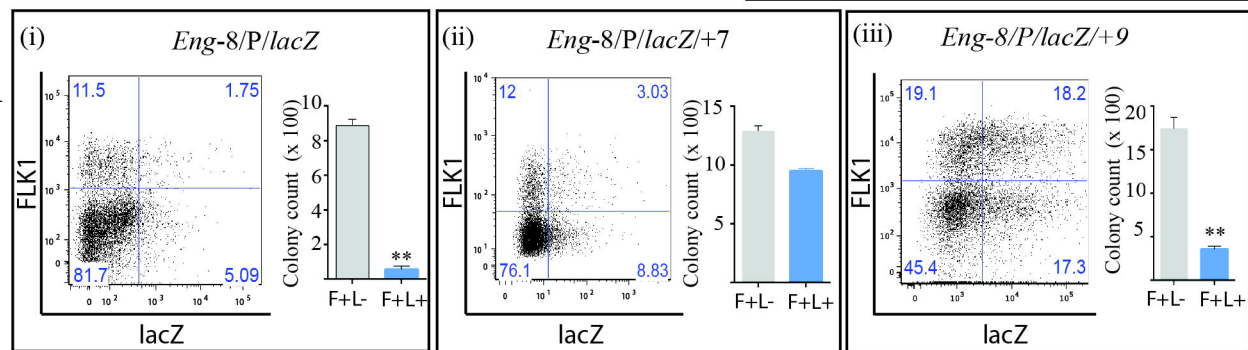
A



B



D



C

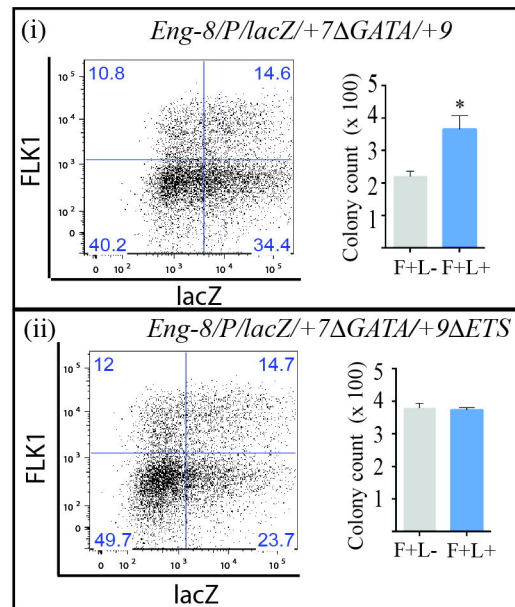


Figure 2

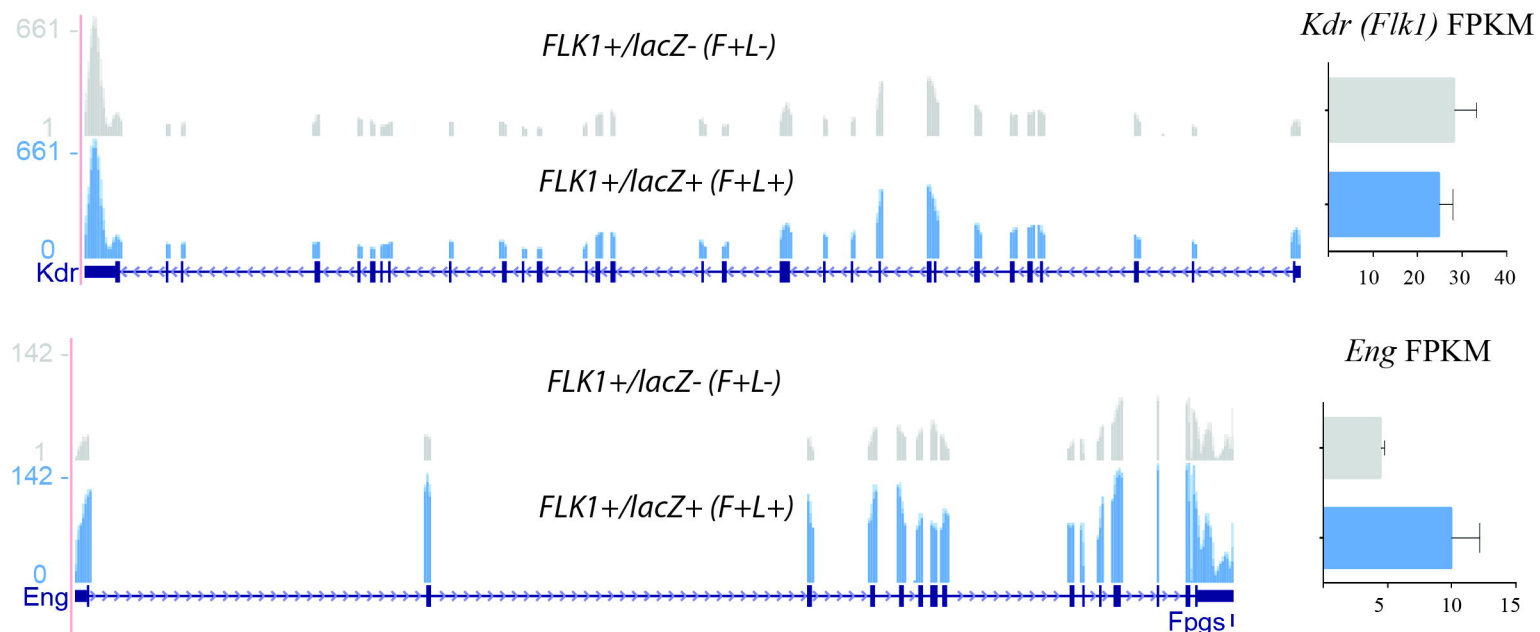
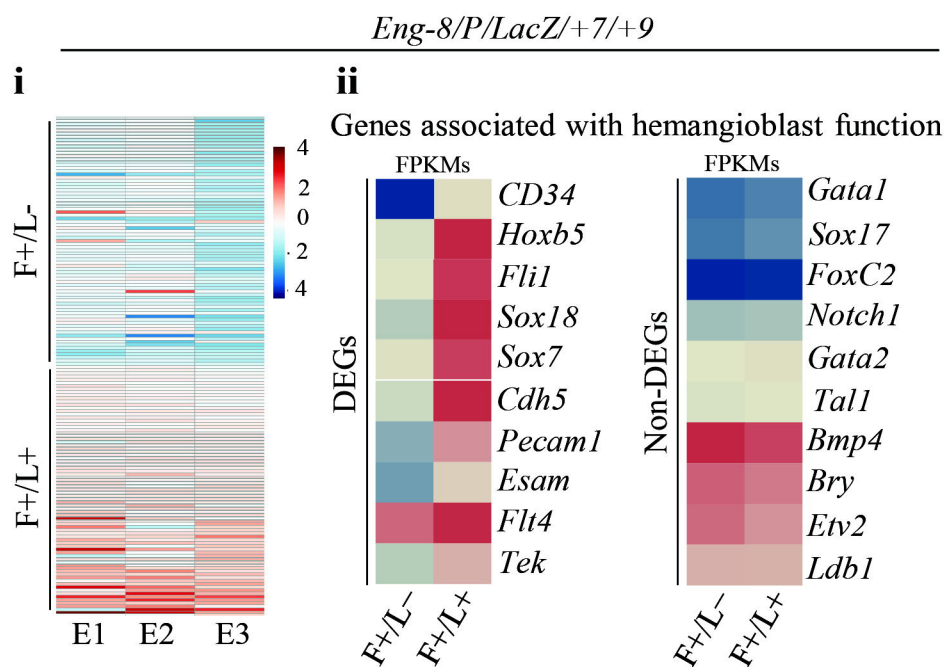
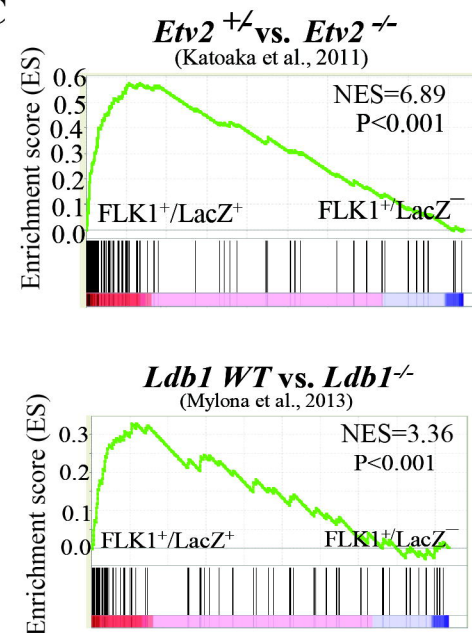
A**B****C**

Figure 3

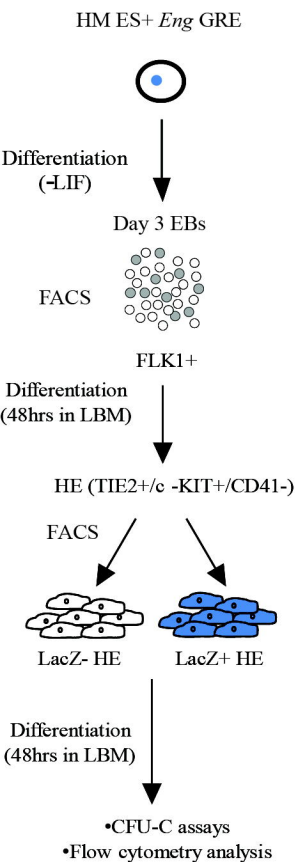
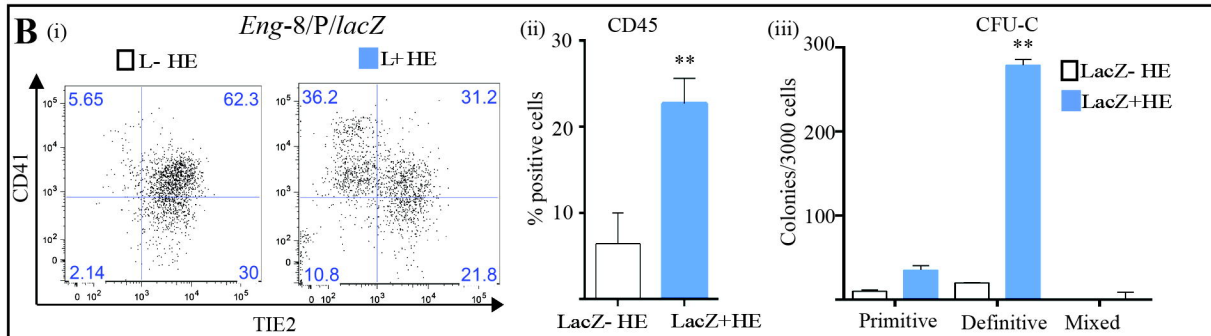
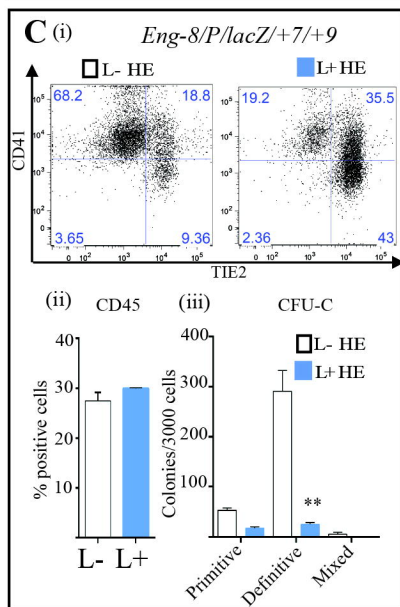
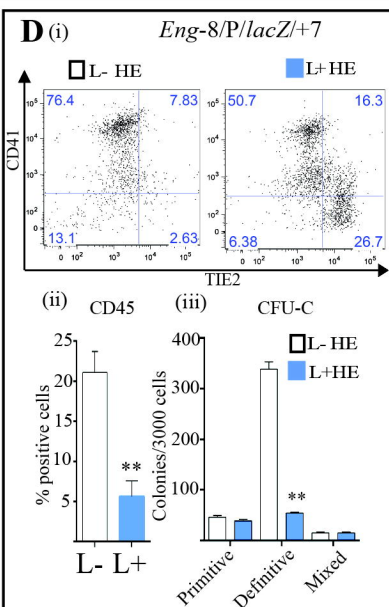
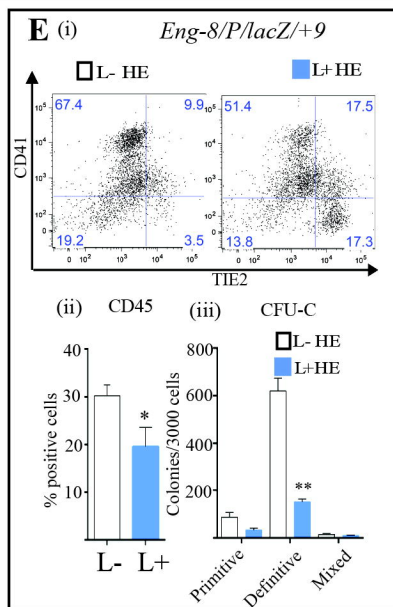
A**B****C****D****E**

Figure 4

```

graph TD
    A[HM ES+ Eng GRE] --> B[Day 3 EBs]
    B --> C[FLK1+]
    C --> D[HE TIE2+/c-KIT+/CD41-]
    D --> E1[ENG+/lacZ- HE]
    D --> E2[ENG+/lacZ+ HE]
    D --> E3[ENG-/lacZ+ HE]
    E2 --> F[Flowcytometry]
    E2 --> G[CFU-C assays]
    E2 --> H[RNA-sequencing]
  
```

HM ES+ Eng GRE

Differentiation (-LIF)

Day 3 EBs

FACS

FLK1+

Differentiation (48hrs in LBM)

HE (TIE2+/c-KIT+/CD41-)

FACS

ENG+/lacZ- HE

ENG+/lacZ+ HE

ENG-/lacZ+ HE

Differentiation (48hrs in LBM)

- CFU-C assays
- RNA-sequencing

- Flowcytometry
- CFU-C assays

Figure 3 consists of three panels. Panel (i) is a flow cytometry plot of lacZ vs ENG for CD41-ve cells, showing three populations with percentages 3.88, 21.5, and 17.5. Panel (ii) shows three flow cytometry plots of c-Kit vs TIE2 for CD41-/E+/L-, CD41-/E+/L+, and CD41-/E-/L+ populations, with percentages 46.6, 40.4, and 14.5 respectively. Panel (iii) is a bar graph of CFU-C colonies per 3000 cells for Primitive, Definitive, and Mixed lineages across E+L-, E+L+, and E-L+ populations. Significance levels are indicated by asterisks.

Figure 3 consists of three panels. Panel (i) shows three flow cytometry plots of c-Kit (y-axis) vs TIE2 (x-axis) for CD41⁻/E⁺/L⁻, CD41⁻/E⁺/L⁺, and CD41⁻/E⁻/L⁺ groups. The plots show the percentage of cells in different quadrants. Panel (ii) is a bar graph showing the percentage of positive cells for CD45 in the E+L-, E+L+, and E-L+ groups. Panel (iii) is a bar graph showing the number of colonies/3000 cells for Primitive, Definitive, and Mixed stages in the E+L-, E+L+, and E-L+ groups. Statistical significance is indicated by asterisks (***) and (**).

Panel	Group	Quadrant 1 (%)	Quadrant 2 (%)	Quadrant 3 (%)	Quadrant 4 (%)	% Positive Cells	Colonies/3000 cells	
(i)	CD41 ⁻ /E ⁺ /L ⁻	5.05	57.1	3.55	34.3	E+L- (green) E+L+ (cyan) E-L+ (red)	Primitive Definitive Mixed	
		CD41 ⁻ /E ⁺ /L ⁺	9.1	58.3	3.11			29.5
	CD41 ⁻ /E ⁻ /L ⁺		46.9	25.9	8.8			18.4
		(ii)	E+L-					~4.5
	E+L+				~3.5			
	E-L+				~19.5			
(iii)	E+L-	~10	~10	~900	~10			
	E+L+	~10	~10	~10	~10			
	E-L+	~10	~10	~10	~10			

Figure 5

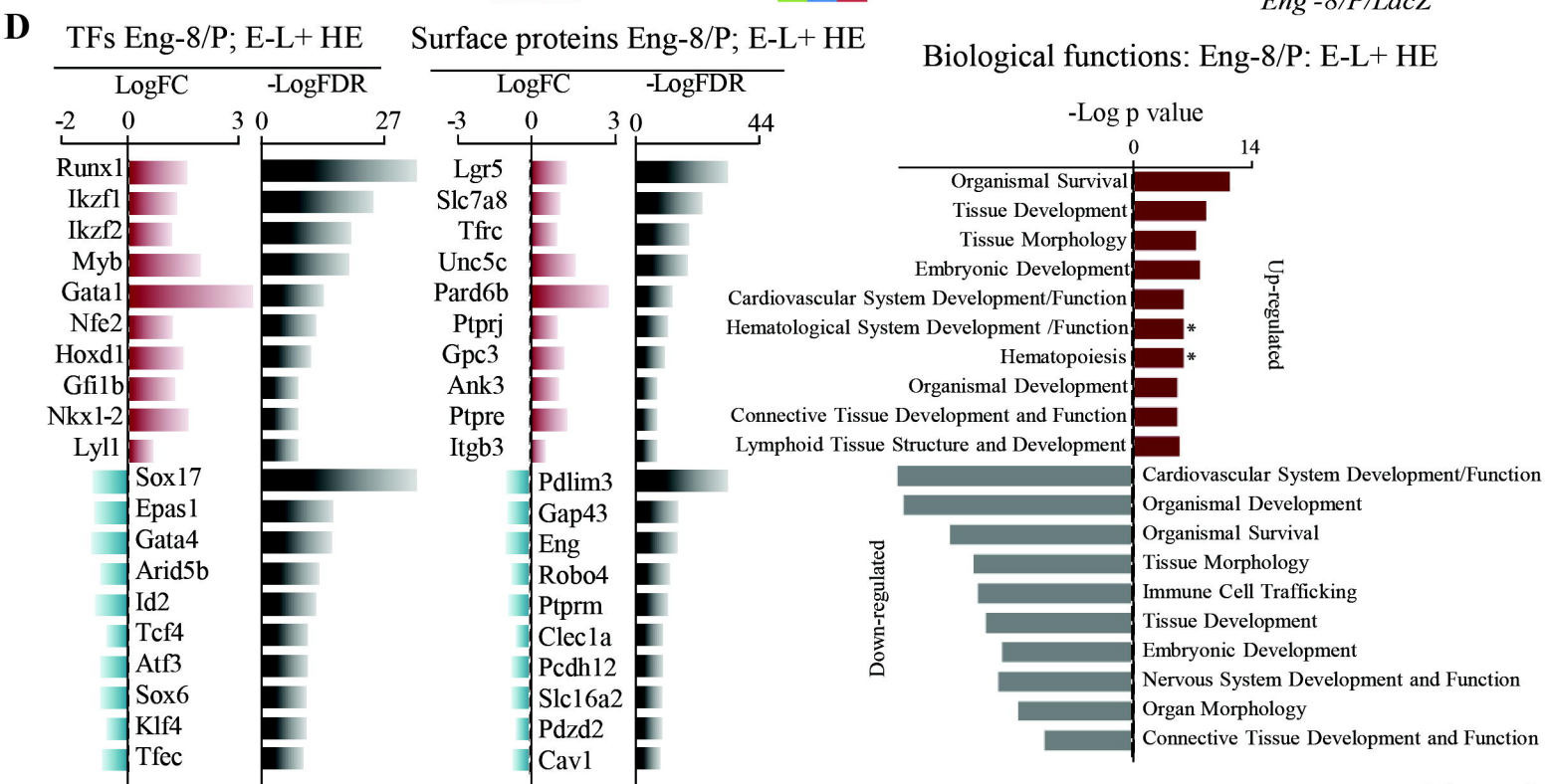
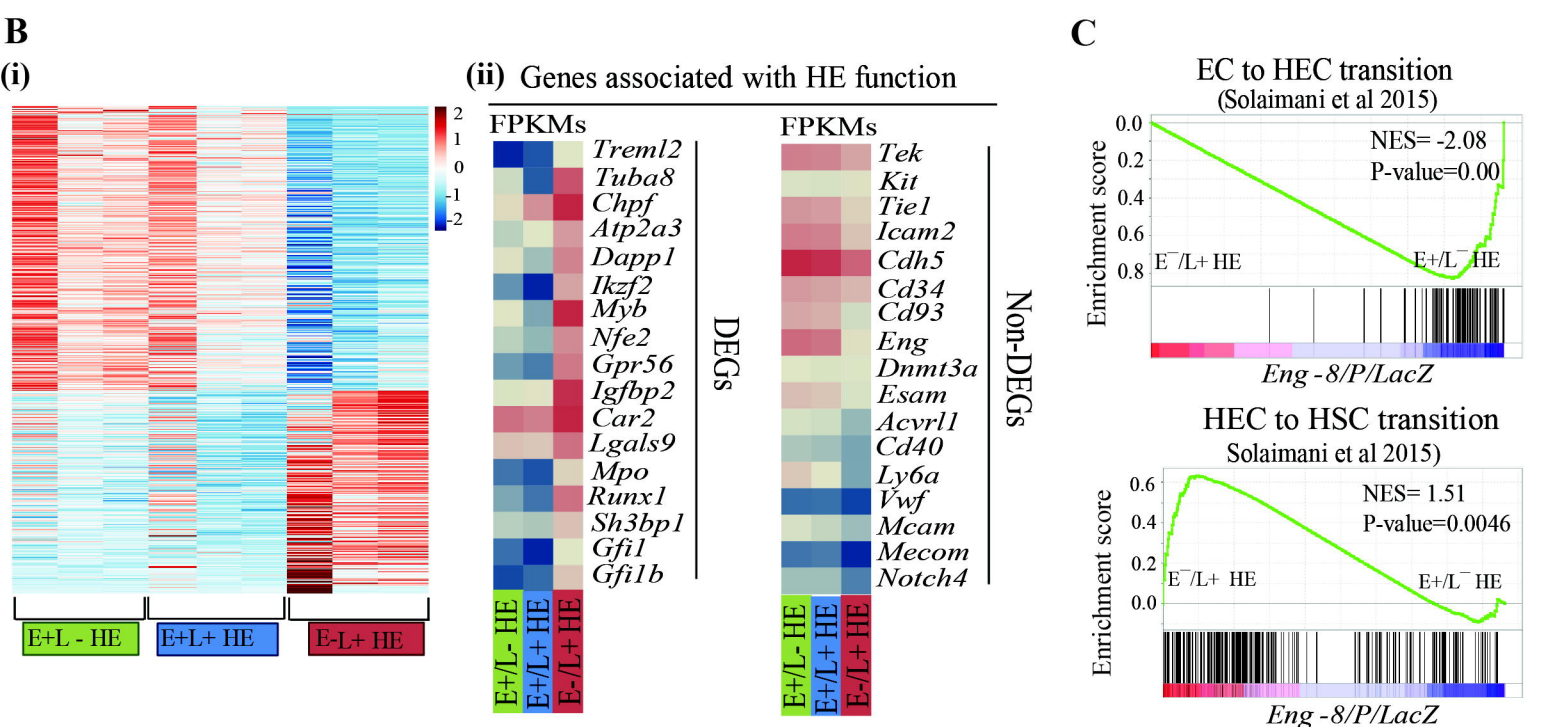
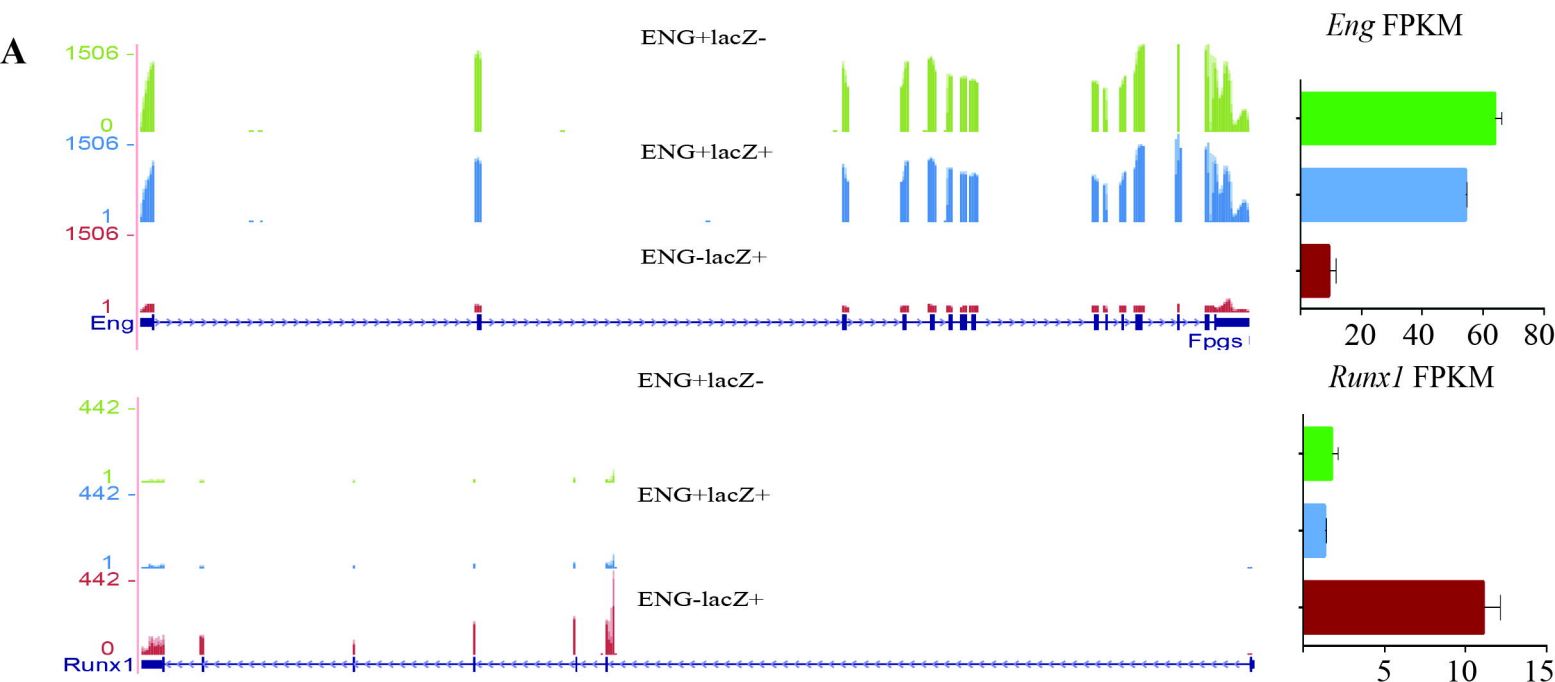


Figure 6

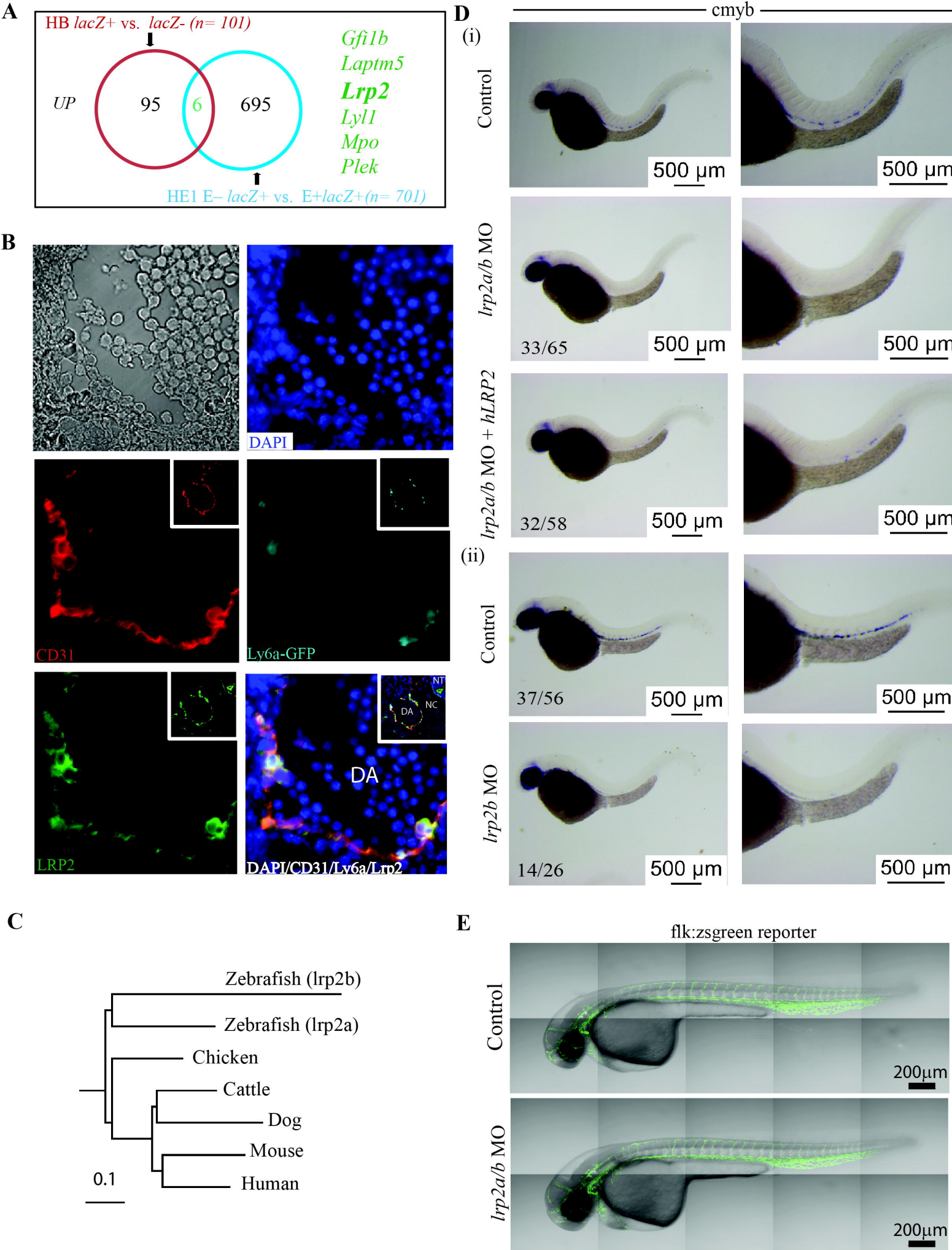


Figure 7